

# Prediction of Sec-dependent secreted proteins in *Rickettsia conorii* (Malish 7) using bioinformatics tools

**Joshua M. Shallom**<sup>1</sup>, Nataraj V. Dongre<sup>1</sup>, Oswald R. Crasta<sup>1</sup>  
Joseph J. Gillespie<sup>1,2</sup>, Abdu Azad<sup>2</sup>, Joao C. Setubal<sup>1</sup> and Bruno  
S. Sobral<sup>1</sup>

<sup>1</sup>Virginia Bioinformatics Institute (VBI), Virginia  
Tech University, VA and <sup>2</sup>University of Maryland,  
Baltimore City, MD



http://patric.vbi.vt.edu

Submit Feedback

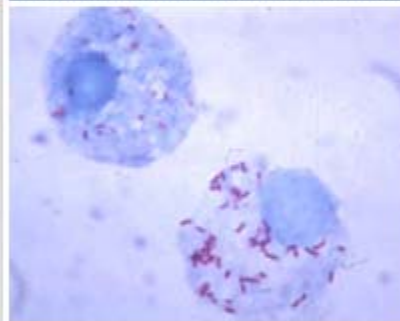
Organisms Searches & Tools Downloads About PATRIC Contact Us

No Items in Cart

## News

- 17 August 2007**  
PATRIC Data Release and Website Update.
- 01 June 2007**  
PATRIC Website Update. This update contains significant changes to the PATRIC website. The changes were designed to make the website easier to use and get the information that you are looking for.
- 16 April 2007**  
PATRIC Data Release and Website Update. This data release contains new genomes for Coronaviruses (18 genomes), Hepatitis A (2 genomes), and Lyssaviruses (14 genomes).
- 15 January 2007**  
PATRIC Data Release and Website Update. This release contains newly curated genomes and website enhancements. Updated curation includes standardization of protein product names and names of mature peptides for all Calicivirus genomes and curation of 10 additional Hepatitis A genomes.

## The Organisms We Study



### Bacteria

- Brucella*
- Coxiella*
- Rickettsia*

### Viruses

- Caliciviridae*
- Coronavirus*
- Hepatitis A Virus*
- Hepatitis E Virus*
- Lyssavirus*

## Assemble & Download Sequences of Interest

- Perform 9 specialized searches
- Use custom filters to drill down and narrow search results in feature tables
- Save and download results with feature cart
- Manipulate orthologous groups of related proteins for all PATRIC organisms

## Perform Comparative Genomics

- Use tools to pinpoint common or unique traits among genes/proteins
- Compare amino acid sequences across proteins using Multiple Sequence Alignment

## Perform Comparative Pathway Analysis

- Compare PATRIC bacterial pathogen genomes at the pathway level



Submit Feedback

Organisms

Searches & Tools

Downloads

About PATRIC

Contact Us

No Items in Cart

## About PATRIC

- What is PATRIC?
- Data Releases & Updates
- Metrics on the PATRIC site
- Standard Operating Procedures
- Personnel
- Outreach & Publications
- Related Sites
- Contact Us

## Personnel

VBI

**Organism Experts**

Scientific Working Group

### Bacteria

#### **Dr. Abdu Azad**

The University of Maryland

*Rickettsia*

#### **Dr. Stephen Boyle**

Virginia-Maryland Regional College of Veterinary Medicine

*Brucella*

#### **Dr. Robert A. Heinzen**

NIAID, NIH

*Coxiella*

#### **Dr. James E. Samuel**

Texas A&M University

*Coxiella*

### Viruses

#### **Dr. Susan Baker**

Loyola University Medical Center

Coronaviruses

#### **Dr. Yuri Khudyakov**

Centers for Disease Control and Prevention

Hepatitis A viruses


#### **Dr. Xiang-Jin Meng**

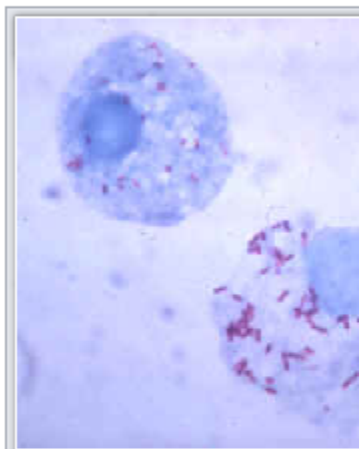
Virginia-Maryland Regional College of Veterinary Medicine

#### **Dr. Charles Rupprecht**

Centers for Disease Control and Prevention

- Genomes
- Proteins
- Genes
- Phylogenetic Tree
- Ortholog Groups
- Pathways
- Additional Resources
- Collaborative Research

Genomes (13)	Accession	Type	Size (bp)
<a href="#">Rickettsia typhi str. Wilmington</a> 	<a href="#">NC_006142</a>	chromosome	1,111,496
<a href="#">Rickettsia prowazekii str. Madrid E</a>	<a href="#">NC_000963</a>	chromosome	1,111,523
<a href="#">Rickettsia sibirica 246</a>	<a href="#">NZ_AABW01000001</a>	chromosome	1,250,021
<a href="#">Rickettsia rickettsii</a>	<a href="#">NZ_AADJ01000001</a>	chromosome	1,257,710
<a href="#">Rickettsia felis URR.WXCal2</a>	<a href="#">NC_007109</a>	chromosome	1,485,148
	<a href="#">NC_007110</a>	plasmid	62,829
	<a href="#">NC_007111</a>	plasmid	39,263
<a href="#">Rickettsia africae ESF-5</a>	<a href="#">NZ_AAUY01000001</a>	chromosome	1,276,710
<a href="#">Rickettsia bellii OSU 85-389</a>	<a href="#">NZ_AARC01000001</a>	chromosome	1,528,980
<a href="#">Rickettsia massiliae MTU5</a>	<a href="#">NZ_AAVR01000001</a>	chromosome	1,360,898
<a href="#">Rickettsia canadensis str. McKiel</a>	<a href="#">NZ_AAFF01000001</a>	chromosome	1,159,772
<a href="#">Rickettsia monacensis</a>	<a href="#">EF564599</a>	plasmid	23,486
<a href="#">Rickettsia bellii RML369-C</a>	<a href="#">NC_007940</a>	chromosome	1,522,076
<a href="#">Rickettsia akari str. Hartford</a>	<a href="#">NZ_AAFE01000001</a>	chromosome	1,231,060
<a href="#">Rickettsia conorii Malish 7</a>	<a href="#">NC_003103</a>	chromosome	1,268,755



**Image**  
Gimenez stain of tick hemolymph infected with *R. rickettsii*. (Courtesy of CDC)

Rickettsiae are members of the alpha-group of the purple bacteria. Along with the families Bartonellaceae and Anaplasmataceae, they are members of the order Rickettsiales and the family Rickettsiaceae. The tribe Rickettsiini now has two genera, *Rickettsia* and *Orientia*. The genus *Rickettsia* is divided into spotted fever and typhus groups. The typhus group consists of *R. prowazekii*, *R. typhi*, *R. canada* and *R. felis* (Sexton et al. 1998).

For references and expanded information, see the [Genome Report](#).

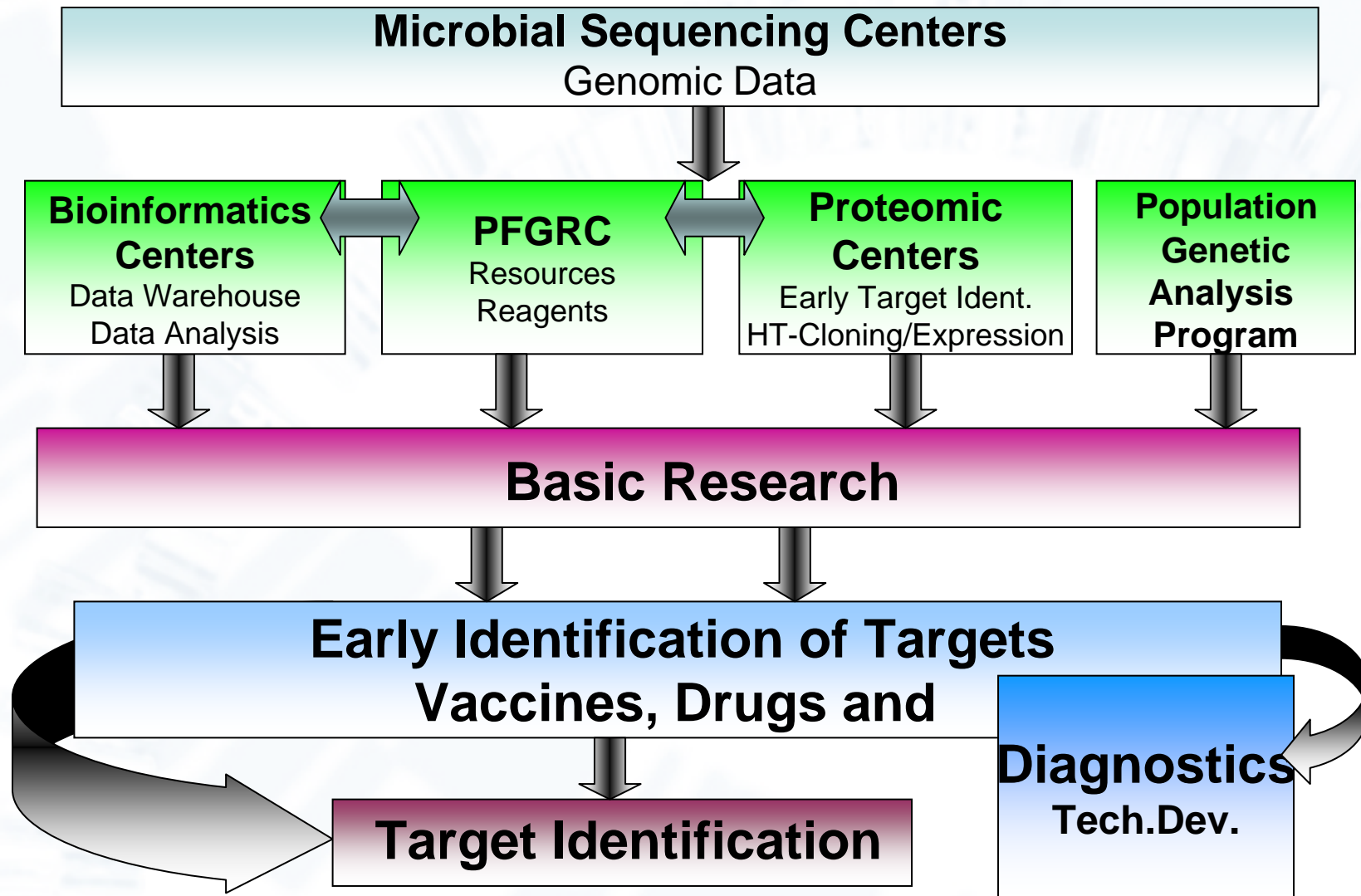
Orthologs	
Ortholog Groups	1955
Ortholog Groups with Annotations	1339
Ortholog Groups with Hypothetical Proteins	616

Using the Orthologous groups for Rickettsia, Joseph Gillespie has done a comparative phylogenomics analysis

See **Poster # 120** later this afternoon

Comparative Phylogenomics of Rickettsia  
Joseph J. Gillespie, Abdu Azad *et. al.*

# NIAID Genomics Program



Source: Valentina Di Francesco, BRC Program Officer, NIAID

# Motivation/Scope

- Identify drug targets/vaccine candidates
- Bacterial surface proteins
- Secreted proteins
- Protein function not defined
- Lab. based prediction slow & expensive
- Programs not perfect

# Why secreted proteins?

- Cell Wall assembly
- Nutrient uptake
- Virulence
- Antibiotic resistance
- Pili and flagella biogenesis
- Adherence
- Energy generation
- Environmental sensing

# Secreted Proteins

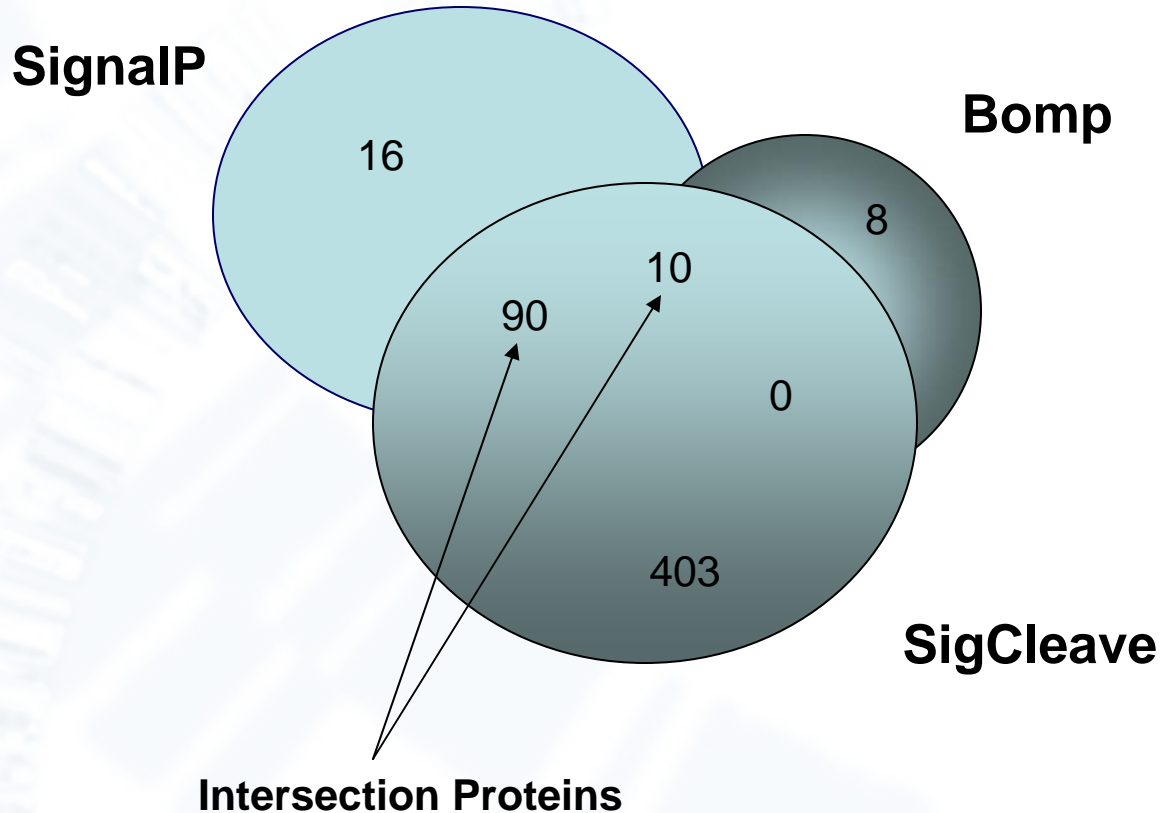
## Methods and Programs

- Input Data: *Rickettsia conorii* genome, 1567 genes. Genome taken after curation at VBI
- SignalP program: Predicts the presence of a signal peptide
- BOMP: Predicts the presence of beta-barrel structures
- SigCleave: Predicts signal peptide cleavage site
- PSortB: Predicts subcellular localization for Gram-negative bacteria

# Choice and Performance values

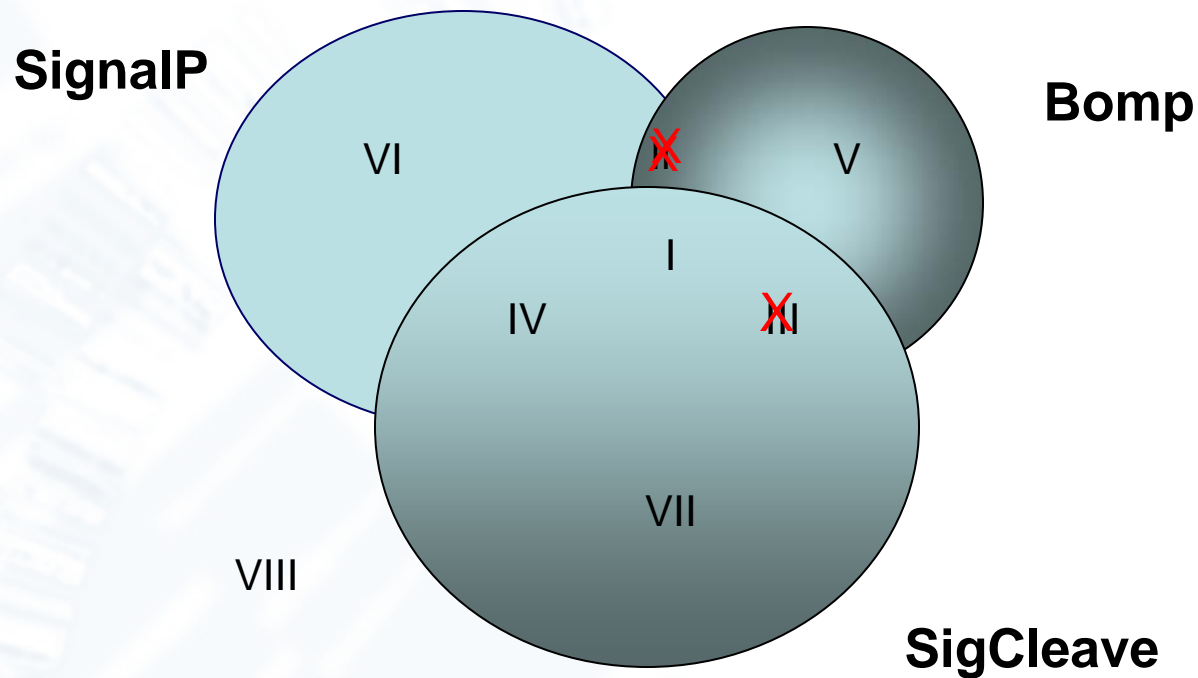
- SignalP - 86 percent
- SigCleave - 67.9 percent
- Bomp - 73.9 percent
- PSortB - C, CM, OM, P, EC  
(Various values)

# Secreted Proteins Results: *R. conorii*



Not to scale

# Confidence Levels (Subjective)



# Confidence Levels (Subjective)

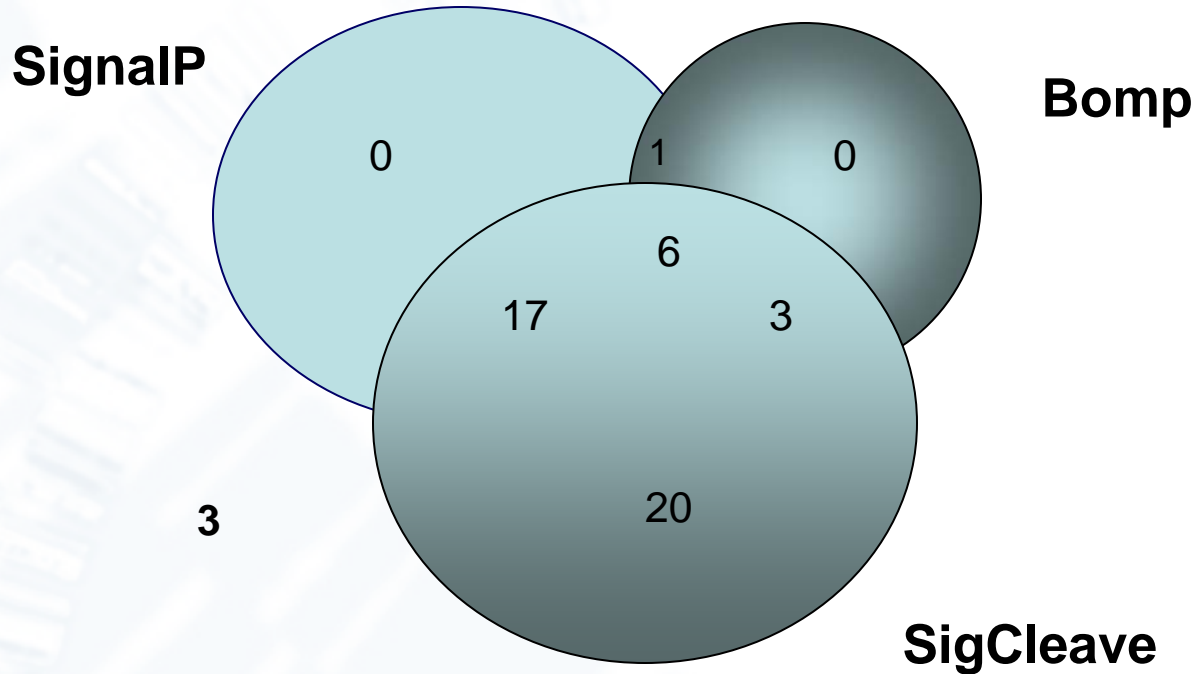
- I – All 3 Programs (SigP/SigCleave/Bomp)
- II – (SigP/Bomp) – SigCleave
- III – (SigCleave/Bomp) – SigP
- IV – (SigP/SigCleave) – Bomp
- V – Bomp – (SigP/SigCleave)
- VI – SigP – (SigCleave/Bomp)
- VII – SigCleave – (SigP/Bomp)
- VIII – None of the Programs (Outside)

# InterProScan results for predicted *R. conorii* secreted proteins

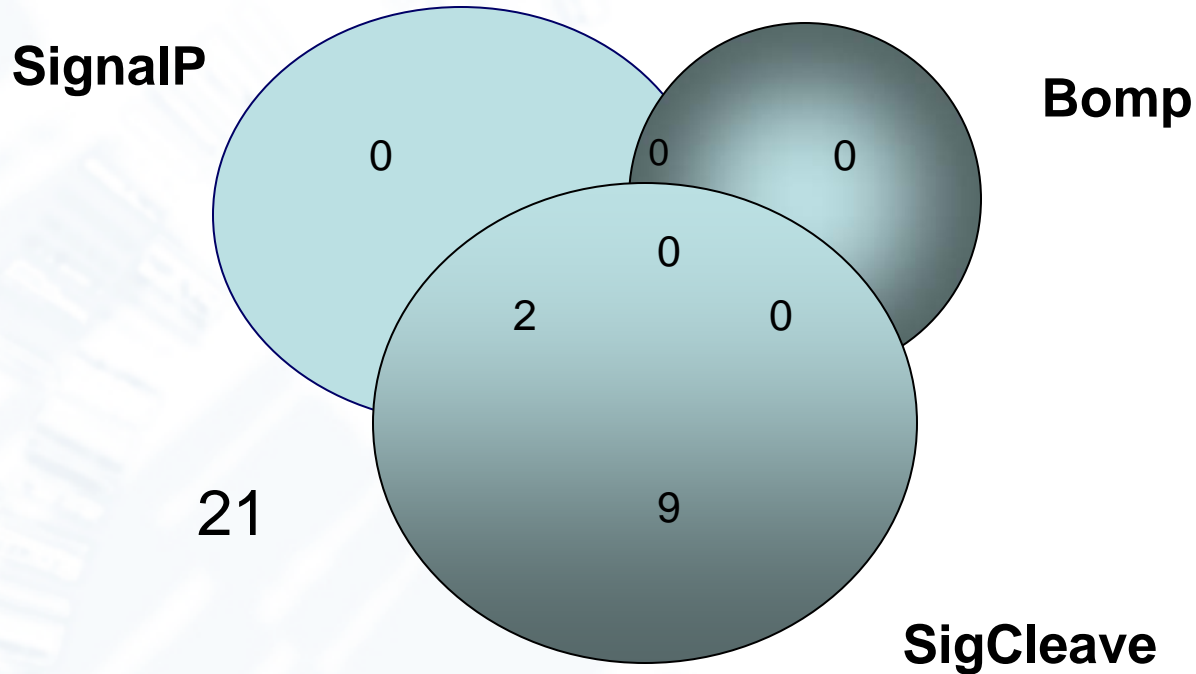
- VirB-6 like Type IV secretion proteins
- Outer membrane proteins
- Peptidase motifs
- Trypsin-like protease motifs
- Organic solvent tolerance protein
- Osmotic shock protein
- Sec secretion system homologs
- Polysaccharide Lyase family protein
- Sca surface antigen proteins
- HlyD family secretion protein
- Autotransporters

**Next step: experimental validation**

# Secreted Proteins Results: Positive Control



# Secreted Proteins Results: Negative Control



# Performance of our method

$$TP = (47/50) \times 100$$

94.0 percent

$$FN = (3/50) \times 100$$

6.0 percent

$$FP = (11/32) \times 100$$

34.4 percent

$$TN = (21/32) \times 100$$

65.6 percent

# Secreted Proteins Future Directions

- Data is available at PATRIC website
- These methods will also be available at PATRIC website
- SwissProt annotation will be prioritized for these proteins
- Collaborations are being sought for wet-laboratory characterization and expression studies

# Acknowledgements

## Project members:

Nataraj Dongre

Joseph Gillespie (UMD)

Abdu Azad (UMD)

Oswald Crasta

Joao Setubal

Bruno Sobral

## Funding:

NIAID/NIH contract number

HHSN26620040035C

awarded to Bruno Sobral

# Virginia Bioinformatics Institute

*September 2007*



Thank you!

Questions?